Metabolic Engineering Communications 3 (2016) 153–163

Contents lists available at ScienceDirect

Metabolic Engineering Communications

journal homepage: www.elsevier.com/locate/mec

Improving the flux distributions simulated with genome-scale metabolic models of *Saccharomyces cerevisiae*

Rui Pereira [a,b], Jens Nielsen [b], Isabel Rocha [a,*]

[a] CEB-Centre of Biological Engineering, University of Minho, Campus de Gualtar, Braga 4710–057, Portugal
[b] Department of Biology and Biological Engineering, Chalmers University of Technology, SE 412 96 Gothenburg, Sweden

ARTICLE INFO

Article history:
Received 7 October 2015
Received in revised form
17 March 2016
Accepted 10 May 2016
Available online 13 May 2016

Keywords:
Genome-scale metabolic model
Saccharomyces cerevisiae
Flux distribution
NADH: NADPH
Metabolic engineering

ABSTRACT

Genome-scale metabolic models (GEMs) can be used to evaluate genotype-phenotype relationships and their application to microbial strain engineering is increasing in popularity. Some of the algorithms used to simulate the phenotypes of mutant strains require the determination of a wild-type flux distribution. However, the accuracy of this reference, when calculated with flux balance analysis, has not been studied in detail before.

Here, the wild-type simulations of selected GEMs for *Saccharomyces cerevisiae* have been analysed and most of the models tested predicted erroneous fluxes in central pathways, especially in the pentose phosphate pathway. Since the problematic fluxes were mostly related to areas of the metabolism consuming or producing NADPH/NADH, we have manually curated all reactions including these cofactors by forcing the use of NADPH/$NADP^+$ in anabolic reactions and NADH/NAD^+ for catabolic reactions. The curated models predicted more accurate flux distributions and performed better in the simulation of mutant phenotypes.

1. Introduction

The application of Genome-Scale Metabolic Models (GEMs) to the design and optimization of microbial strains is of great biotechnological value and has been the topic of extensive research (Garcia-Albornoz and Nielsen, 2013; Milne et al., 2009). The interest in this topic has fuelled the development of numerous computational methods that can simulate the fluxes within GEMs and search for combinations of genetic modifications that optimize a desirable phenotypical trait (Kim et al., 2014; Lewis et al., 2012; Xu et al., 2013). These strain engineering methods have been applied to the optimization of the production of a range of different chemicals, which resulted in the construction of microbial strains with improved phenotypes (Asadollahi et al., 2009; Bro et al., 2006; Brochado et al., 2010; Choi et al., 2010; Fowler et al., 2009; Otero et al., 2013).

Although the performance of *in silico* "inspired" strains is better than the wild-type, the experimental production yields are often below the simulated values. One of the factors that can affect the quantitative reproducibility of the results obtained with strain engineering algorithms is the quality of the GEM. If a GEM does not represent the full extent of the metabolic activities of the host organism or contains errors, it might lead to inconsistencies between simulations and *in vivo* data.

GEMs are normally validated by testing their predictions of core physiological parameters, such as the maximum growth yield under different environmental conditions or the set of essential genes in a particular condition (Edwards and Palsson, 2000; Feist et al., 2007; Mo et al., 2009). The validation process rarely includes the analysis of the intracellular fluxes, even for models of well-studied organisms, such as *Escherichia coli* and *Saccharomyces cerevisiae* (Edwards and Palsson, 2000; Feist et al., 2007; Förster et al., 2003; Mo et al., 2009). One of the possible reasons for the lack of validation of GEMs at the flux level might be the existence of multiple objective functions to simulate flux distributions, which can yield different results for the same model and result in a subjective quality assessment.

The accuracy of the fluxes simulated with GEMs is of special relevance when the simulation algorithms used for computing mutant phenotypes require a reference (wild-type) flux distribution, such as MOMA (Segrè et al., 2002), lMOMA (Becker et al., 2007), ROOM (Shlomi et al., 2005) or MiMBL (Brochado et al., 2012). Given the absence of flux data at the genome scale, the reference flux distribution used by the algorithms stated above is usually calculated using Flux Balance Analysis (FBA; Orth et al., 2010) or parsimonious Flux Balance Analysis (pFBA; Lewis et al.,

Abbreviations: GEM, genome-scale metabolic model; pFBA, parsimonious flux balanced analysis; FBA, flux balance analysis; BPCY, biomass product coupled yield; PPP, pentose phosphate pathway

* Corresponding author.
E-mail address: irocha@deb.uminho.pt (I. Rocha).

http://dx.doi.org/10.1016/j.meteno.2016.05.002
2214-0301/

2010) (Alper et al., 2005; Jung et al., 2010; Park et al., 2007). Consequently, given the lack of validation of GEMs at the flux prediction level, the reference flux distributions calculated using FBA/pFBA can contain flux values that are not consistent with *in vivo* values, which may compromise the subsequent phenotype simulations. The sensitivity of phenotype simulations towards the reference fluxes has been demonstrated before for the MOMA algorithm (Segrè et al., 2002), showing that even small deviations in the wild-type flux distribution can have an impact on the simulation outcome. Furthermore, Brochado et al. (2012) also noted that the reference flux distribution had a significant impact when MiMBL was used to simulate single gene deletion mutants. Therefore, it is of special interest to analyse the validity of flux distributions calculated with commonly used methods, such as FBA/pFBA, in order to prevent inaccurate reference flux distributions from affecting negatively the phenotypical predictions for mutant organisms.

Besides the accuracy of the reference flux distribution, there are also other ways of improving the results of phenotype simulation algorithms. For example, there is a group of simulation methods that uses gene expression data to improve the simulation accuracy (Machado and Herrgård, 2014). Methods such as RELATCH (Kim and Reed, 2012) can analyse the changes in gene expression between the wild-type and the mutant strain to infer about the direction and magnitude of flux changes expected. However, these methods require gene expression data, for both the wild-type and mutant organism, which reduces their application to the simulation of strains that are already constructed and characterized at the transcriptomic level.

Fig. 1 summarizes the steps required to perform simulations of mutant phenotypes using algorithms that require a reference set of fluxes. As mentioned above, GEMs are usually not validated at the flux level and not much is known about the accuracy of the reference flux distributions obtained with FBA/pFBA. Therefore, we propose that additional steps should be included in the *in silico* strain engineering framework in order to verify the accuracy of the flux distributions used by the simulation algorithms. When the fluxes obtained are not consistent with the knowledge available for a particular organism, further curation and validation of the model should be performed.

S. cerevisiae is one of the most studied microorganisms and a considerable amount of information about its metabolism is available. It has proven to be quite versatile in terms of industrial applications, making it very attractive for metabolic engineering. Several GEMs have been published for *S. cerevisiae* over the last ten years (Duarte et al., 2004; Förster et al., 2003; Herrgård et al., 2008; Kuepfer et al., 2005; Mo et al., 2009; Nookaew et al., 2008; Österlund et al., 2013) and their predictions of maximum growth yield and exchange fluxes using FBA is generally quite good. These facts make *S. cerevisiae* very attractive for investigating in detail the capacity of GEMs and simulation methods to predict the intracellular flux distribution of the central carbon metabolism.

Here we analysed the flux distributions simulated with pFBA for a selection of *S. cerevisiae* GEMs and verified that there are considerable differences in the predicted fluxes among the models tested. The underlying causes for the differences observed were traced back to the production and consumption of NADPH/NADH, which led us to perform a detailed analysis and curation of all reaction involved in the metabolism of these cofactors. Additionally, we also investigated the impact of using different flux distributions on the simulation of mutant phenotypes. The production of two organic acids was used as case-study to validate if the improved flux distributions affected the prediction of the simulation algorithm. Overall, the improvements made to the metabolism of NADPH resulted in phenotype simulations more consistent with physiological data available in the literature, which supports the assumption that *in silico* strain design could benefit from additional validation steps as here proposed.

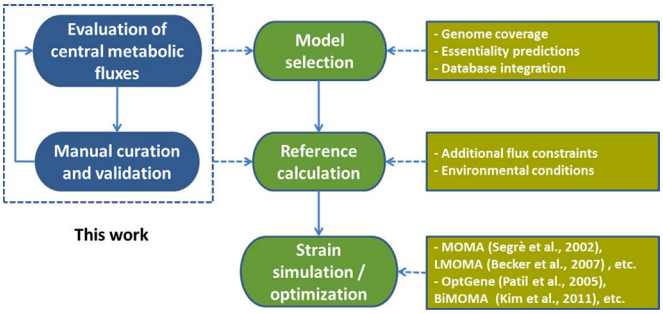

Fig. 1. Contextualization of the work described here within the workflow of *in silico* strain design. Cited algorithms: MOMA (Segrè et al., 2002), LMOMA (Becker et al., 2007), OptGene (Patil et al., 2005), BiMOMA (Kim et al., 2011).

2. Methods

2.1. Model retrieval and pre-processing

The models *i*FF708 (Förster et al., 2003) and *i*TO977 (Österlund et al., 2013) were obtained in SBML format from the authors' website (http://biomet-toolbox.org/index.php? page=models-S.cerevisiae). The model *i*MM904 (Mo et al., 2009) was downloaded and converted to SBML using the script provided by the authors in the supplementary material. Yeast 6.06 (Heavner et al., 2013) was downloaded in SBML format from the project's website: http://sourceforge.net/projects/yeast/files/.

Each model was imported into OptFlux 3.07 (Rocha et al., 2010) and the *in silico* environmental conditions were set to mimic minimal growth media supplemented with glucose under fully-aerobic conditions (ammonia: unconstrained uptake, phosphate: unconstrained uptake, sulfate: unconstrained uptake, oxygen: unconstrained uptake, glucose: fixed uptake of 1.15 mmol/gCDW h).

2.2. Flux simulations

All simulations were executed within OptFlux 3.07 using IBM ILOG CPLEX optimization studio (academic) as the linear programming solver. The simulation method chosen to compute the flux distributions was pFBA (Lewis et al., 2010) and the objective function was defined as the maximization of the biomass production.

In order to compare experimental flux data with the simulations it was necessary to create an equivalence map between the main metabolic reactions present in central carbon metabolism and the reactions of the models. Each model was inspected manually to determine which reaction (or reactions) corresponds to an enzymatic activity from the central metabolism. In genome-scale metabolic models it is common to find duplicated reactions and other stoichiometrically equivalent pathways that catalyse the same chemical transformation. The phosphorylation of glucose, for example, can be carried by hexokinases or glucokinases. To simplify the visualization of the fluxes through the central metabolism of *S. cerevisiae,* the fluxes through equivalent reactions were summed into a single value. Using the example above, all the reactions phosphorylating glucose to glucose-6-phosphate present in each model were treated as a single entity. The equivalence map for each of the models tested is provided in supplementary tables S6 and S7.

2.3. Flux variability analysis

To verify the uniqueness of the flux distributions calculated by FBA, each reaction's variability was tested by flux variability analysis (Burgard et al., 2001; Mahadevan and Schilling, 2003). Briefly, the value of the objective function (biomass growth) was fixed at its maximum and each target flux was then minimized or maximized. However, since the models tested here contained many reactions in the central metabolism that are duplicated, when one of them is minimized the flux could be routed through the other. Furthermore, if the duplicated reactions are reversible, futile cycles can arise. To avoid these irrelevant glitches in the analysis, when one reaction's variability was tested, all the other "equivalent partners" were disabled.

2.4. Optimizations of metabolic engineering targets

In order to search for strain designs with improved production of acetate or mevalonate, the evolutionary optimization algorithms included in OptFlux were used (Patil et al., 2005; Rocha et al., 2008). The method selected to simulate the mutant phenotypes was the linear version (Becker et al., 2007) of the MOMA algorithm (Segrè et al., 2002), where the Euclidian distance of the fluxes is replaced by the Manhattan distance. Two different phenotype objective functions were optimized by the evolutionary algorithm: the Biomass Product Coupled Yield (BPCY; Patil et al., 2005) or the maximization of the target flux assuming a minimum biomass growth threshold (YIELD). The evolutionary algorithm was run at least three times for each target, setting the number of solution evaluations to 50,000 and the maximum number of knock-outs allowed to 5. In order to reduce the total number of possible knock-out targets, the set of essential reactions and dead-end reactions was not considered in the optimizations.

3. Results and discussion

3.1. Comparison of genome-scale metabolic models

Since the first GEM of *S. cerevisiae* was published in 2003 (Förster et al., 2003), several authors have expanded and improved the metabolic reconstruction of this microorganism (reviewed by Nookaew et al. (2011) and Osterlund et al. (2012)). Table 1 shows the details for a selection of the GEMs available in the literature for *S. cerevisiae*, focusing on model size and other relevant parameters. The analysis of Table 1 reveals that over the years additional ORFs and metabolic reactions were progressively included in the models, which resulted in models of increasing size. In comparison to *i*FF708, the newer models include additional compartments and some of them were balanced at the elemental level.

The validation of a GEM is usually performed using large-scale gene essentiality data and/or extracellular flux predictions (reviewed by Sánchez and Nielsen, 2015). Since the capability of these models to predict the intracellular flux distributions is rarely analysed, it is hard to select the most suitable model for applications that require the computation of a reference set of fluxes (such as MOMA, ROOM, etc.). In order to clarify the suitability of the GEMs of *S. cerevisiae* to predict the internal flux distribution of this microorganism, we selected four different models from Table 1 (*i*FF708, *i*MM904, *i*TO977 and Yeast 6) and used them for simulations with pFBA under the same environmental conditions. *i*FF708 was chosen for this analysis because it is still widely used for phenotype simulation purposes (Asadollahi et al., 2009; Bro et al., 2006; Brochado et al., 2010; Otero et al., 2013), and the three other models were chosen as they are the three most recent ones (Table 1). The flux values were taken from a pFBA simulation and normalized to a glucose uptake rate of 100 arbitrary units.

When the models selected for analysis were simulated under the same conditions it was noticeable that Yeast 6 was predicting a maximum specific growth rate considerably higher than the others (0.16 h^{-1} vs. 0.11 h^{-1}). A detailed analysis of this problem revealed that an unreasonably high flux through the ATP synthase was generating ATP without a corresponding input of NADH in the oxidative phosphorylation pathway. This odd behaviour was caused by a cycle of proton export from the mitochondria, originating from the aspartate:proton symporter (reaction r_1117 in the model). By using a series of transporters and metabolic reactions, Yeast 6 was capable of creating a proton gradient across the mitochondrial membrane without any energy input. The solution to this problem was to disable the activity of the aspartate:proton symporter in the direction from the mitochondria to the cytosol. Since under aerobic conditions there is a proton gradient across the mitochondria membrane, the transport of aspartate is more likely to occur in the direction of this gradient, i.e., from the cytosol to the mitochondria. This issue allowed the ATP synthase to be active even in the absence of oxygen, which can also explain why it is necessary to manually disable this enzyme to obtain valid anaerobic simulations with this model (Heavner et al., 2013). The problem with the unconstrained transport of protons from the mitochondria to the cytosol in the model Yeast 6 has been patched in the most recent version of the Yeast consensus model (see revision 7.11 in http://yeast.sourceforge.net/).

Using the amended Yeast 6 and the other models the flux distributions were computed and represented in a metabolic map of the central metabolic pathways as shown in Fig. 2. One of the most noticeable differences observed is the absence of flux in the oxidative Pentose Phosphate Pathway (PPP) in the most recent models (*i*MM904, *i*TO977 and Yeast 6) in comparison to *i*FF708. An analysis of the fluxes revealed that in *i*TO977 and *i*MM904 all the NADPH required for anabolic processes originated from the cytosolic NADP-specific isocitrate dehydrogenase, which is inconsistent with the role of the PPP in the production of this cofactor (it is estimated that 26–44% of the total glucose that is consumed enters the PPP [Gombert et al., 2001; Jouhten et al., 2008]). Regarding the Yeast 6 model, we observed that the main reaction fixing ammonia was the NADH-dependent glutamate

Table 1
Details of selected GEMs available for *S. cerevisiae*.

Model	Year	ORFs included	Reactions included	Other information
*i*FF708 (Förster et al., 2003)	2003	708	1175	3 compartments (2 metabolic[a])
*i*ND750 (Duarte et al., 2004)	2004	750	1489	8 compartments (7 metabolic), Elementally balanced
*i*LL672 (Kuepfer et al., 2005)	2005	672	1038	3 compartments (2 metabolic)
*i*IN800 (Nookaew et al., 2008)	2008	800	1446	3 compartments (2 metabolic)
*i*MM904 (Mo et al., 2009)	2009	904	1412	8 compartments (7 metabolic), Elementally balanced
*i*TO977 (Österlund et al., 2013)	2013	977	1566	4 compartments (3 metabolic)
Yeast 6 (Heavner et al., 2013)	2013	900	1888	15 compartments (7 metabolic), Elementally balanced

[a] Metabolic compartments refer to aqueous intracellular compartments, while the non-metabolic include the extracellular space and cellular membranes.

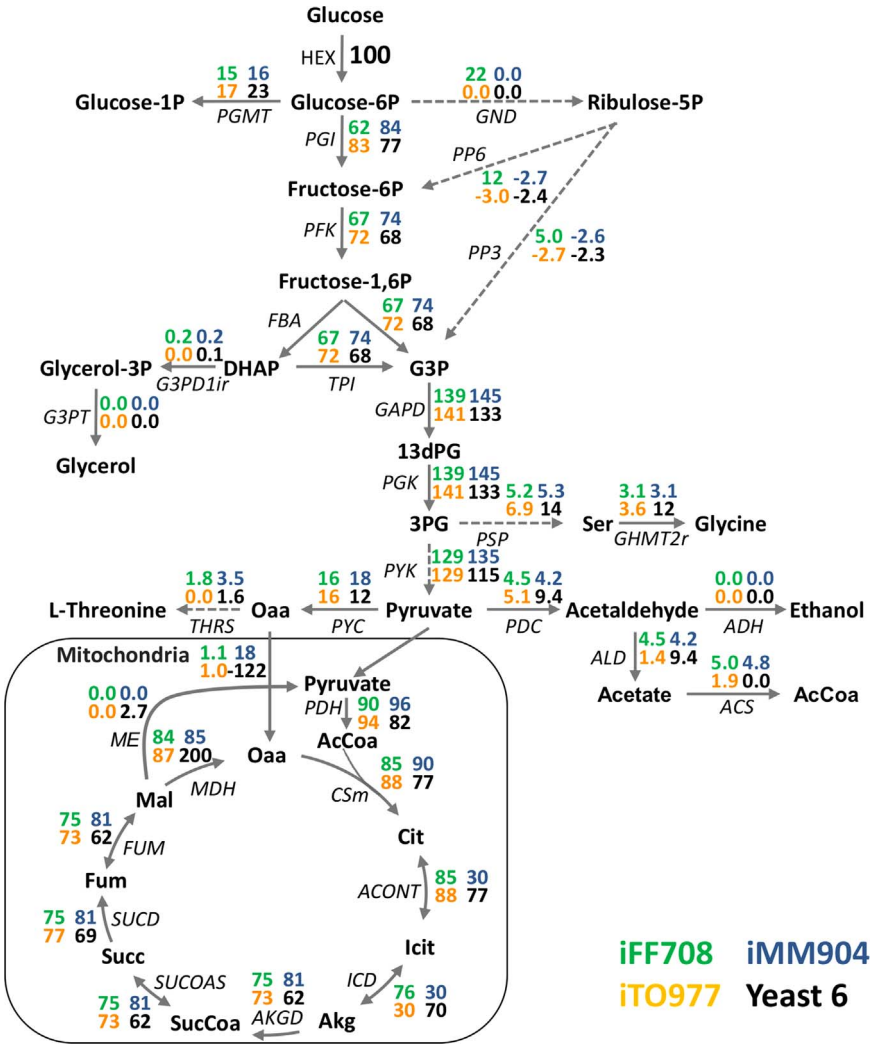

Fig. 2. Distribution of fluxes in the central metabolism predicted by pFBA for the models: *i*FF708 (green), *i*MM904 (blue), *i*TO977 (orange) and Yeast 6 (black). The flux values shown are normalized to a glucose uptake rate of 100 arbitrary units. Growth rates for each model: *i*FF708–0.11 h^{-1}, *i*MM904–0.11 h^{-1}, *i*TO977–0.11 h^{-1}, Yeast 6–0.09 h^{-1}. Reactions: ACONT- aconitase, ACS- acetyl-CoA *syn*thetase, ADH- alcohol dehydrogenase, AKGD- alpha-ketoglutarate dehydrogenase, ALD- aldehyde dehydrogenase, CSm- citrate synthase, FBA- fructose 1,6-bisphosphate aldolase, FUM- fumarase, G3PD1ir- glycerol-3-phosphate dehydrogenase, G3PT- glycerol-1-phosphatase, GAPD- glyceraldehyde-3-phosphate dehydrogenase, GHMT2r- serine hydroxymethyltransferase, GND- 6-phosphogluconate dehydrogenase, HEX- hexokinase, ICD- mitochondrial isocitrate dehydrogenase, MDH- mitochondrial malate dehydrogenase, ME- mitochondrial malic enzyme, PDC- pyruvate decarboxylase, PDH- pyruvate dehydrogenase, PFK- phosphofructokinase, PGI- phosphoglucose isomerase, PGK- 3-phosphoglycerate kinase, PGMT- phosphoglucomutase, PP3- sum of the non-oxidative reactions of the pentose phosphate pathway producing glyceraldehyde-3-phosphate, PP6- sum of the non-oxidative reactions of the pentose phosphate pathway producing fructose-6-phosphate, PSP- phosphoserine phosphatase, PYC- pyruvate carboxylase, PYK- pyruvate kinase, SUCD- succinate dehydrogenase, SUCOAS- succinyl-CoA ligase, THRS- threonine synthase, TPI- triose phosphate isomerase. Metabolites: 13dPG- 1,3-diphosphoglycerate, 3PG- 3-phosphoglycerate, AcCoA- acetyl-CoA, Akg- 2-oxoglutarate, Cit- citrate, DHAP- dihydroxyacetone-phosphate, Fum- fumarate, G3P- glyceraldehyde-3-phosphate, Icit- isocitrate, Mal - L-malate, Oaa- oxaloacetate, Ser- L-serine, Succ- succinate, SucCoa- succinyl-CoA.(For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

dehydrogenase instead of the NADPH dependent variant (flux not shown). Furthermore, in Yeast 6, the NADPH required for most anabolic processes was being produced by the cytosolic aldehyde dehydrogenase and the cytosolic C1-tetrahydrofolate synthase.

The model comparison shown in Fig. 2 indicates that among the tested models *i*FF708 shows the distribution of fluxes most consistent with flux distributions of *S. cerevisiae* in glucose limited chemostats (Gombert et al., 2001; Jouhten et al., 2008). In comparison to *i*FF708, the remaining models performed worst mostly in the fluxes related to NADPH metabolism. Therefore, in order to improve the flux accuracy of the most recent *S. cerevisiae* GEMs it seems imperative to curate the areas of the metabolism involved in NADPH production and consumption.

3.2. Model curation based on cofactor utilization

Given all the indications pointing towards problems with the metabolism of NADPH in the models of *S. cerevisiae* under analysis, all the metabolic reactions involving this cofactor were collected and their feasibility was examined. Since NADH can replace NADPH as the electron donor in many reactions of *S. cerevisiae*, the reactions where NADH was present were also included in the list of reactions to be curated.

The curation process started by analysing one of the major issues detected in the course of the model comparison, which was the production of NADPH by the cytosolic isocitrate dehydrogenase in the models *i*MM904 and *i*TO977. The gene coding for the cytosolic isocitrate dehydrogenase (*IDP2*) is known to be repressed by glucose (Haselbeck and McAlister-Henn, 1993), but in glucose-limited conditions, its activity might still be present. Therefore, one cannot rule out the contribution of this gene to the overall NADPH production based on regulatory phenomena alone. The work of Satrustegui et al. (1983) has shown that the NADPH/$NADP^+$ ratio can also act as an inhibition factor for the cytosolic

isocitrate dehydrogenase. In the same publication, Satrustegui et al. (1983) reported that under growth on glucose the NADPH/$NADP^+$ ratio is high enough to inhibit the cytosolic isocitrate dehydrogenase of *S. cerevisiae*. In addition to this inhibitory effect, it is also relevant to point out that the concentration of NADPH/$NADP^+$ can also affect the change in Gibbs free energy of this reaction, which should favour the equilibrium in the direction of NADPH consumption.

Assuming that the findings of Satrustegui et al. (1983) can be generalized to other enzymes that use NADPH or NADH as cofactors we curated manually all cytosolic reactions containing either of these cofactors in the models under analysis. Since under fully-aerobic glucose-limited conditions the ratio of NADPH/$NADP^+$ is usually high, in order to drive anabolic reactions, and the ratio of NADH/NAD^+ is usually low, in order to promote fast metabolic oxidation of the substrate (Canelas et al., 2008; Dijken and Scheffers, 1986; Satrustegui et al., 1983; Voet and Voet, 2011), the reversibility of cytosolic reactions should favour the consumption of NADPH and the production of NADH. Therefore, most of the reactions present in each model were constrained in the direction of NADPH consumption and NADH production to enforce cofactor availability constraints. The curation process is documented in the supplementary Tables S1–S4 and the final list of constrained reactions is shown in Table 2. The constraints from Table 2 are not intended to be used as general revisions of the models tested here, but rather as condition specific flux bounds. Since the cofactor abundance assumed in the curation process might not be valid for other cultivation conditions, the flux constraints should be used exclusively for simulating wild-type *S. cerevisiae* in glucose limited fully oxidative conditions.

During the manual curation process we observed that some of the flux constraints applied to enforce cofactor availability on the *S. cerevisiae* models (Table 2) resulted in a lethal phenotype. The inability to produce some biomass components after applying the cofactor constraints was analysed in order to determine if the observed phenotype was a consequence of a faulty model structure or misleading assumptions about the reaction's feasibility. For *i*FF708 two constraints suffered from this issue: the inactivation of the squalene epoxidase (model ID: ERG1) and the inactivation of the cytosolic isocitrate dehydrogenase (model ID: IDP2_1). In the first case, ERG1 was inactivated because it could produce NADPH. However, this reaction is described to consume NADPH (Leber et al., 1998), supported by the fact that the other models all had the correct stoichiometry. Therefore, we corrected the stoichiometry of ERG1 reaction in the model *i*FF708 and removed it from the list of reactions to be inactivated. Regarding the lethal phenotype associated with the inactivation of the cytosolic isocitrate dehydrogenase in *i*FF708, the reason for this behaviour was traced back to the lack of mitochondrial transporters for 2-oxoglutarate. Using the available knowledge about *S. cerevisiae* metabolism, we added two transporters for 2-oxoglutarate to *i*FF708: citrate/2-oxoglutarate antiporter (*YHM2*) and malate/2-oxoglutarate antiporter (*ODC1, ODC2*) (Castegna et al., 2010; Palmieri et al., 2001).

In the case of *i*MM904 a similar issue was encountered when the cytosolic isocitrate dehydrogenase (model ID: R_ICDHy) was inactivated. As before, the addition of the 2-oxoglutarate transporters solved this issue and restored normal growth. However, following the addition of these transporters there was an abnormal increase in biomass formation caused by an artificial cycle of proton export from the mitochondria, similar to the one observed for the original Yeast 6 and discussed in Section 3.1. It was solved by changing the reversibility of three mitochondrial transporters: aspartate:proton symporter (model ID: R_ASPt2m), oxaloacetate:proton symporter (model ID: R_OAAt2m) and mitochondrial dicarboxylate carrier for L-malate (model ID: R_MALtm) were allowed only to transport into the mitochondria. These changes

Table 2
List of changes in the flux bounds applied during the curation process for each model (reactions are identified by their IDs on the model). The reversibility constrained category includes all reversible reactions that were made irreversible to impose cofactor availability constraints and the inactivated category includes all irreversible reactions that were inactivated to impose cofactor availability constraints.

Model	Reversibility constrained		Inactivated	
	NADH	NADPH	NADH	NADPH
***i*FF708**	MDH3, MDH2, ADH2, ADH1, ADH5, ADH4, BIO2, SFA1_2, SFA1_1, LYS1, TDH1, TDH2, TDH3	HMG2, HMG1, ECM17, MET10, LYS9	U45_, PRO3_3, FAS1_4, FOX2, PRO2_1, LYS2_2, HOM6_1	TYR1, IDP3_1, ARA1_2, ALD6, IDP2_1
***i*MM904**	BTDD-RR, FALDH, 2HBO, ALCD2x, SACCD2, MDH, GAPD	HMGCOAR, SACCD1, SULR	AASAD2, G5SD2, HPROa, ALCD25xi, ALCD2ir, ALCD26xi, ALCD22xi, FMNRx, ALCD23xi, C22STDSx, LNS14DMx, HSDxi, ALCD24xi	LSERDHr, ATHRDHr, PPND2, GLYCDy, ARAB1D2, ICDHy, ALDD20y, ALDD2y
***i*TO977**	BIO2, LYS1, r275, ADH1, SFA1_1, MDH2	HMG1, FAS2_1_2, FAS2_2_2, FAS2_3_2, FAS2_4_2, FAS2_5_2, FAS2_6_2, FAS2_7_2, FAS2_8_2, FAS1_1_2, FAS1_2_2, FAS1_3_2, FAS1_4_2, FAS1_5_2, FAS1_6_2, FAS1_7_2, FAS1_8_2, LYS9, ECM17	LYS2_2, PRO2_1, U45_, PRO3_3, HOM6_1, PGA3	ARA1_2, r581, r572, r671, IDP2_1, r253, r833, ALD6, TYR1
Yeast 6.06	r_0714, r_0486, r_0470, r_0003	–	r_0169, r_0166, r_0179, r_0182, r_0186, r_0441, r_1010, r_2115	r_0173, r_0177, r_0234, r_0659, r_0676, r_0690, r_0321, r_0939

A

Reaction	Original models				Curated models				Experimental	
	iFF708	iMM904	iTO977	Yeast 6	iFF708	iMM904	iTO977	Yeast 6	Jouhten et al., (2008)	Gombert et al. (2001)
Glycolysis / PPP										
PGI	62	84	83	77	52	49	53	49	49	34
GND	22	0.0	0.0	0.0	32	36	30	29	27	44
TAL/TKT1	7.4	-0.1	-0.3	-0.1	11	12	9.6	9.5	8.5	14.6
TKT2	5.0	-2.6	-2.7	-2.3	8.3	9.3	7.2	7.5	6.5	12.1
Other										
ICDHc	8.2	60	53	0.0	0.0	0.0	0.0	0.0	-	-
ICDHm	76	30	30	70	81	79	75	80	71	60
GDH1	43	46	47	0.0	43	45	47	38	-	-
GDH2	0.0	0.0	0.0	49	0.0	0.0	0.0	0.0	-	-
ALD	4.5	4.2	1.4	9.4	4.5	4.2	1.4	2.9	7.0	46
OAAt	1.1	18	1.0	-122	5.9	2.1	1.0	4.6	16	18

B

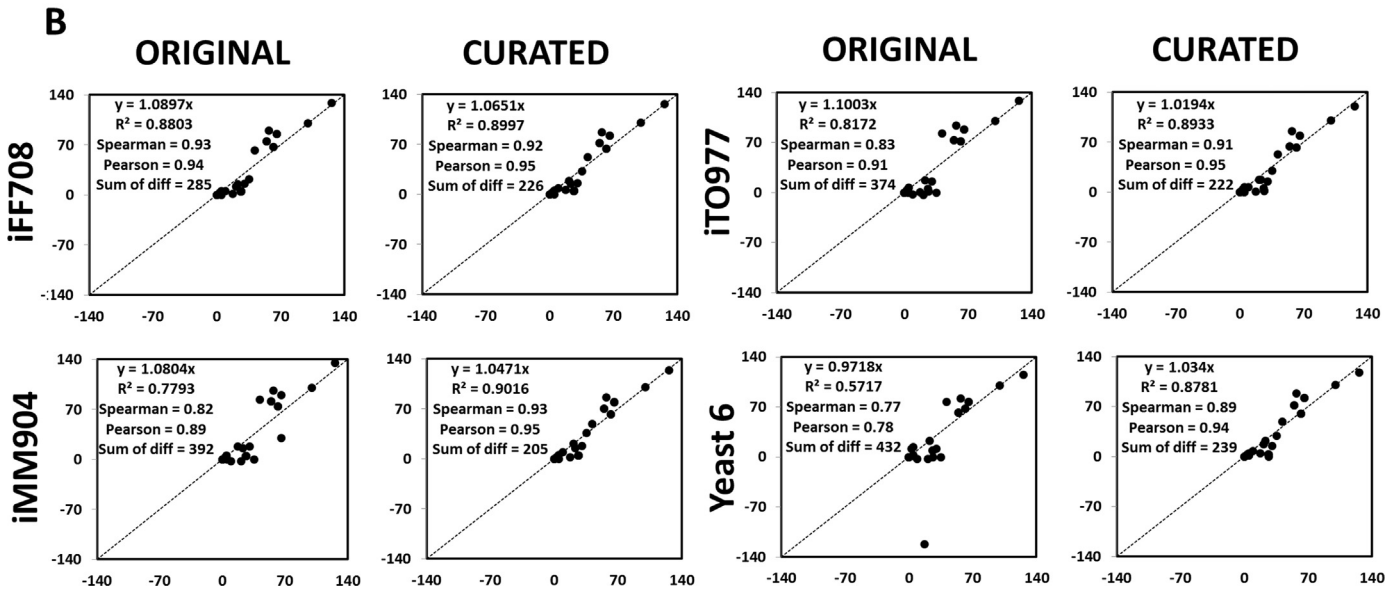

Fig. 3. Comparison between the original and curated flux distributions for the models iFF708, iMM904, iTO977 and Yeast 6. A- Most relevant changes in the flux distributions of the curated models simulated with pFBA in comparison to two experimental flux data sets obtained with 13C-Metabolic Flux Analysis of labelled glucose chemostats (Gombert et al., 2001; Jouhten et al., 2008); B- comparison between the experimentally determined fluxes (*x*-axis) and pFBA simulation of each model (*y*-axis). For each graph it is shown: the linear regression equation, the correlation coefficient (R^2), the Spearman correlation coefficient, the Pearson correlation coefficient and the sum of differences. The experimental fluxes used in the correlations were obtained by averaging both sets of data (Gombert et al., 2001; Jouhten et al., 2008). The flux values shown are normalized to a glucose uptake rate of 100 arbitrary units. Abbreviations: ALD- aldehyde dehydrogenase, GDH1- $NADP^+$-dependent glutamate dehydrogenase, GDH2- NAD^+-dependent glutamate dehydrogenase, GND- 6-phosphogluconate dehydrogenase, ICDHc- cytosolic NADP-specific isocitrate dehydrogenase, ICDHm- mitochondrial isocitrate dehydrogenase, OAAt- oxaloacetate transport from the cytosol to the mitochondria, PGI- phosphoglucose isomerase, TAL- Transaldolase, TKT1- transketolase (D-xylulose-5P + ribose-5P <-> sedoheptulose-7P+glyceraldehyde-3P), TKT2- transketolase (D-erythrose-4P+D-xylulose-5P <-> D-fructose-6P+glyceraldehyde-3P).

restored normal growth levels.

Regarding Yeast 6, the list of reactions to be inactivated also included two candidates that were essential for growth: sterol dehydrogenase (model ID: r_0234) and prephenate dehydrogenase (model ID: r_0939). In both cases, these reactions were present in Yeast 6 as $NADP^+$ dependent, but information from the literature pointed to NAD^+ dependent versions (Baudry et al., 2001; Mannhaupt et al., 1989). Thus, the alternative NAD^+ versions of both enzymes were introduced into Yeast 6, which solved the lethal phenotypes observed.

Using the GEMs containing the flux restrictions from Table 2 and the corrections described above, we simulated again the flux distributions using pFBA and analysed the differences between the curated models and the original ones (Fig. 3(A)). To serve as reference values we also included in Fig. 3(A) two sets of experimentally determined flux values obtained with ^{13}C-Metabolic Flux Analysis of labelled glucose chemostats (Gombert et al., 2001; Jouhten et al., 2008). Fig. 3(A) shows that the curation process resulted in an increase in the oxidative PPP flux (GND) in all models, which was more dramatic in iMM904, iTO977 and Yeast 6. Furthermore, there were also other improvements in reactions related to NADPH/NADH metabolism. The complete set of central

metabolic fluxes is shown in supplementary Fig. S1 for each model under analysis.

Focusing the analysis on *i*FF708 (Fig. 3(A)), there was an increase in the flux through the oxidative PPP to levels closer to the experimental fluxes. This increase was a consequence of an extra consumption of NADPH for anabolic purposes. Since many of the anabolic reactions present in the original *i*FF708 can consume either NADH or NADPH, when the NADH consuming partners were constrained (Table 2) the cell had to synthesize essential biomass components using NADPH as an electron donor. As a consequence, the flux through the PPP had to compensate for the increased requirements for this cofactor. Another visible difference is the increased transport of oxaloacetate into to the mitochondria (OAAt), which can be explained by the inactivation of the cytosolic isocitrate dehydrogenase (ICDHc) and the addition of the mitochondrial 2-oxoglutarate transporters. These modifications shifted 2-oxoglutarate production from the cytosol to the mitochondria, resulting in an increase in the anaplerotic flux of oxaloacetate into the mitochondria.

Regarding *i*MM904, one of the most noticeable problems observed in the model comparison (Fig. 2) was the lack of flux in the oxidative PPP. As shown in Fig. 3(A), in the curated *i*MM904 the flux through the oxidative PPP (GND) increased dramatically in comparison to the original model and the simulated flux value is within the interval of values determined experimentally (Gombert et al., 2001; Jouhten et al., 2008). The modification that contributed the most to the improvement observed in the oxidative PPP was the deactivation of the cytosolic isocitrate dehydrogenase (ICDHc), which shifted NADPH production to the PPP. As a consequence of the increase in flux in the oxidative PPP, the fluxes in the non-oxidative PPP (TAL, TKT1 and TKT2) changed direction when compared to simulations of the original model and became consistent with the experimental flux data. There was also an improvement in the mitochondrial isocitrate dehydrogenase (ICDHm) as a consequence of the deactivation of the cytosolic isocitrate dehydrogenase in conjunction with the introduction of the 2-oxoglutarate transporters.

Similarly to *i*MM904, *i*TO977 also showed improvements in PPP and citric acid cycle. In this case, the reason for the zero flux through the PPP was solely the activity of the cytosolic isocitrate dehydrogenase (ICDHc), which was supplying all the NADPH required for growth. Furthermore, this issue was also responsible for the low flux through mitochondrial isocitrate dehydrogenase (ICDHm). Consequently, the cofactor constrains applied to the model solved both of these problems (Fig. 3(A)).

In the case of Yeast 6, the curated model showed improvements on several pathways, which included the PPP, transport of oxaloacetate and in the ammonium assimilation. Similarly to the other models, the increase in the oxidative part of the PPP (GND) is one of the most prominent differences in comparison to the original model simulation. Additionally, the non-oxidative reactions of the PPP (TAL, TKT1 and TKT2) changed direction in comparison to the original model and became consistent with experimental flux values. Another considerable difference between the original and curated models is the cofactor usage by the glutamate dehydrogenase reaction. After the curation process, the cofactor used in the glutamate dehydrogenase reaction changed from NADH (GDH2) to NADPH (GDH1), which increased the requirements for NADPH considerably. The preference of *S. cerevisiae* for the NADPH dependent glutamate dehydrogenase is consistent with the importance of this enzyme *in vivo* (Avendaño et al., 1997; Dubois et al., 1974). We also observed an improvement in the prediction of the flux of oxaloacetate into the mitochondria in the curated model (OAAt) and a decrease in the cytosolic aldehyde dehydrogenase (ALD), which was the source of NADPH in the simulations made for the original model.

In order to analyse quantitatively the improvements observed in the curated models, we plotted the simulated fluxes against the corresponding experimental values in Fig. 3(B) and analysed statistically the correlation between them. For the revised *i*FF708 model all the parameters analysed improved slightly with the exception of the Spearman coefficient. The modest improvements observed for the curated *i*FF708 are consistent with the fact this model had the best initial flux distribution. However, for the other three models the improvements are much more pronounced, and in all cases the correlation coefficients show that the curation process resulted in a much better prediction of metabolic fluxes in the central metabolism.

Given the non-uniqueness of the FBA/pFBA simulations, it is possible that the flux distributions computed with these methods suffer from variability even when the objective function is at its optimal value. Therefore, the flux changes analysed in this section could have been a consequence of flux variability and not related with the curation process undertaken. In order to test this, we performed Flux Variability Analysis (Burgard et al., 2001; Mahadevan and Schilling, 2003) for each flux shown in Fig. 2. The fluxes that showed more than 10% variation in comparison to the value obtained with pFBA are listed in supplementary Table S5. The analysis of the variable flux values showed that the changes observed between the original and curated models were a consequence of the constraints applied. A thorough analysis of the variable flux values is given in supplementary Text S1.

The genome-scale models modified in this study are provided in supplementary material in SBML format. Two versions of each model are supplied: one version includes general corrections of the problems we encountered in each model during the curation process and another version with all the cofactor constrains shown in Table 2.

3.3. Additional flux inconsistencies

After the manual curation process, some of the fluxes in the central carbon metabolism were still incorrectly simulated by pFBA. One of such cases was the absence of flux through the normal pathway for L-threonine production in the model *i*TO977 (THRS in supplementary Fig. S1). The absence of flux in this pathway was caused by the existence of an alternative route for L-threonine production using a threonine aldolase reaction that converts glycine and acetaldehyde to L-threonine in this model (model ID: GLY1). Although the threonine aldolase reaction is present in all models, only in *i*TO977 it can occur in the direction of L-threonine production. This reaction is reported in the literature as a possible source of glycine in *S. cerevisiae* (McNeil et al., 1994; Monschau et al., 2006), but it has not been reported yet that it can also act as a source of L-threonine. In fact, *S. cerevisiae* becomes auxotrophic for L-threonine when threonine synthase is inactivated (Kingsbury and McCusker, 2010; Mannhaupt et al., 1990), which suggests that the threonine aldolase cannot act as an L-threonine source in this organism.

In the model Yeast 6 we also detected an issue regarding the synthesis of acetyl-CoA in the cytosol. As shown in supplementary Fig. S1, no flux is predicted through acetyl-CoA *syn*thetase (ACS in supplementary Fig. S1), which seems strange given the essentiality of acetyl-CoA synthesis in the cytoplasm. Again, an erroneous reversibility was responsible for this problem: the cytosolic acetyl-CoA hydrolase (model ID: r_0110) was able to catalyse the reverse reaction, joining acetate and coenzyme A without an energy input. This enzyme is reported to be located in the mitochondria and there is no indication that it can catalyse the ligation of acetate and coenzyme A (Fleck and Brock, 2009).

3.4. Impact of reference flux distributions in the simulation of mutant phenotypes

Among the simulation methods available to compute the phenotype of a mutant organism using GEMs, there is a group which assumes that the mutant cell will try to minimize the magnitude of its adaptation in comparison to the wild-type. Within this group of simulation methods there are several alternatives, such as MOMA (Segrè et al., 2002), lMOMA (Becker et al., 2007), ROOM (Shlomi et al., 2005) or MiMBL (Brochado et al., 2012), which differ in the metric used for minimising the distance between the wild-type and the mutant cell. These simulation algorithms require a reference distribution of fluxes to be used as an indication of the metabolic status in the wild-type organism and a few studies have indicated that the simulation results of MOMA and MiMBL are sensitive to the set of fluxes used as reference (Brochado et al., 2012; Segrè et al., 2002).

Since the simulation methods mentioned above rely on a set of reference fluxes for their computations, their results might be affected by the inaccurate fluxes discussed in Section 3.1. To test how the improvements observed in the curated flux distributions (Section 3.2) would perform in comparison to the original ones, we used the model Yeast 6 and the lMOMA algorithm (Becker et al., 2007) to compute the phenotype of strains obtained for the production of two organic acids, acetate and mevalonate.

In order to obtain strains with increased production *in silico* of acetate and mevalonate we used the evolutionary algorithm included in OptFlux to search for sets of optimal knock-outs. During the optimization process we used the curated flux distribution as a reference but the original reaction constraints were kept unchanged. Since some of the knock-outs implemented during the optimization process could alter the abundance of NADPH or NADH, the flux constraints derived from the curation process (Table 2) might not be applicable to the mutant strains simulated.

Table 3 shows a selection of the strain designs obtained with OptFlux for the optimization of mevalonate and acetate production by using the improved flux distribution as a reference during the strain optimization process. One strain design for each case study is presented in Table 3, with detailed information about the gene knock-outs necessary to achieve the corresponding production yields. The results shown for the original reference were obtained by simulating the same strain designs with the flux distribution for Yeast 6 show in Fig. 2.

As shown in Table 3, the mevalonate producing mutant obtained in the optimization procedure required the inactivation of three different metabolic reactions (corresponding to four genes) to achieve a yield of 0.056 g mevalonate/g glucose. A detailed inspection of the changes in the flux distribution (Fig. 4) revealed that the deletion of the NADPH-dependent glutamate dehydrogenases (*GDH1* and *GDH3*) causes a surplus in the availability of NADPH. The deletion of the NAD^+-dependent 5,10-methylenetetrahydrafolate dehydrogenase (*MTD1*) is necessary to avoid the transfer of electrons from NADPH to NAD^+, forcing the excess of NADPH to be used elsewhere. With the inactivation of the succinyl-CoA synthetase (gene name - *LSC1,* ID in Fig. 4 - SUCOAS) the citric acid cycle is broken and part of the flux is redirected to pyruvate decarboxylase (PDC), which originates an excess of acetyl-CoA that is re-assimilated using the enzymes from glyoxylate cycle. Furthermore, this deletion also causes an excess of acetyl-CoA in the mitochondria, which can be condensed into hydroxymethylglutaryl-CoA (HMGCoa) and transported to the cytosol. In that compartment, this metabolite can be converted to mevalonate while consuming NADPH, which alleviates the excess of this cofactor. In summary, two deletions are required to increase the amount of NADPH available in the cytosol, while the other increases the flux to hydroxymethylglutaryl-CoA (precursor of mevalonate). If the original reference flux distribution would be used in the phenotype simulation, the amount of mevalonate excreted by this mutant would be around seven times lower (Table 3), which shows the impact of the reference distribution of fluxes in phenotype simulation. The differences observed in the mevalonate yield are a consequence of the dependence of the mevalonate production on an engineering strategy based on NADPH availability.

Experimental evidence is available in the literature to support part of the strategy presented here for *in silico* mevalonate production. The inactivation of *GDH1* was implemented by Nissen et al. (2000) to alter cofactor abundance in anaerobic conditions to decrease glycerol accumulation. Furthermore, Scalcinati et al. (2012) also used the *GDH1* deletion to increase the abundance of NADPH and consequently boost the flux through the mevalonate pathway as part of a metabolic engineering strategy to increase α-santalene production. The production of cubebol also benefited from the increased NADPH pool resulting from *GDH1* deletion (Asadollahi et al., 2009). Given the experimental evidences supporting the *in silico* predictions, it seems that the improved reference flux distribution is much better to represent the metabolism of the wild-type *S. cerevisiae*, especially in areas related to NADPH production/consumption.

Regarding the acetate producing mutant, there is also a considerable difference in the acetate yield depending on which reference flux distribution is used (Table 3). A quick analysis of the gene deletions required to induce acetate production revealed that, similarly to the mevalonate producing mutant, NADPH metabolism is also involved in this case. However, in this case the key to produce acetate *in silico* is forcing the cytosolic $NADP^+$-dependent aldehyde dehydrogenase (gene name – *ALD6,* ID in Fig. 4 - ALD) to be the source of cytosolic NADPH (Fig. 4).

To reach the acetate producing phenotype, glucose-6-phosphate dehydrogenase (gene name – *ZWF1,* ID in Fig. 4 - GND) must be deleted so that the oxidative PPP becomes inactivated. Since this pathway is the main source of NADPH in the cytosol, the metabolic network has to adapt to the lack of this cofactor by activating other possible production sources. Furthermore, the list

Table 3
Impact of flux distributions in the performance of strain designs obtained with OptFlux for mevalonate and acetate production using lMOMA (the standard names of the deleted genes are shown in bold).

Target	Inactivated reactions	Product yield (g/g glucose)
Mevalonate	***MTD1***: NAD^+ +5,10-methylene-THF -> NADH +5,10-methenyl-THF ***LSC1*** or ***LSC2***: ADP(m)+ HPO_4^{2-}(m)+ Succinyl-CoA(m) -> CoA(m)+ Succinate(m)+ ATP(m) ***GDH1*** and ***GDH3***: H^+ + NH_4+ NADPH +2-oxoglutarate -> H_2O +L-glutamate + $NADP^+$	Corrected reference: 0.056 Original reference: 0.0075
Acetate	***ZWF1***: D-glucose-6-phosphate + $NADP^+$ -> H^+ + NADPH +6-phosphogluconolactone ***IDP2***: $NADP^+$ + Isocitrate -> CO_2+ NADPH +2-oxoglutarate ***GDH2***: NAD^+ + H_2O +L-glutamate <-> H^+ + NADH + NH_4+2-oxoglutarate ***AAT2***: L-aspartate +2-oxoglutarate <-> L-glutamate + Oxaloacetate	Corrected reference: 0.045 Original reference: 0.0

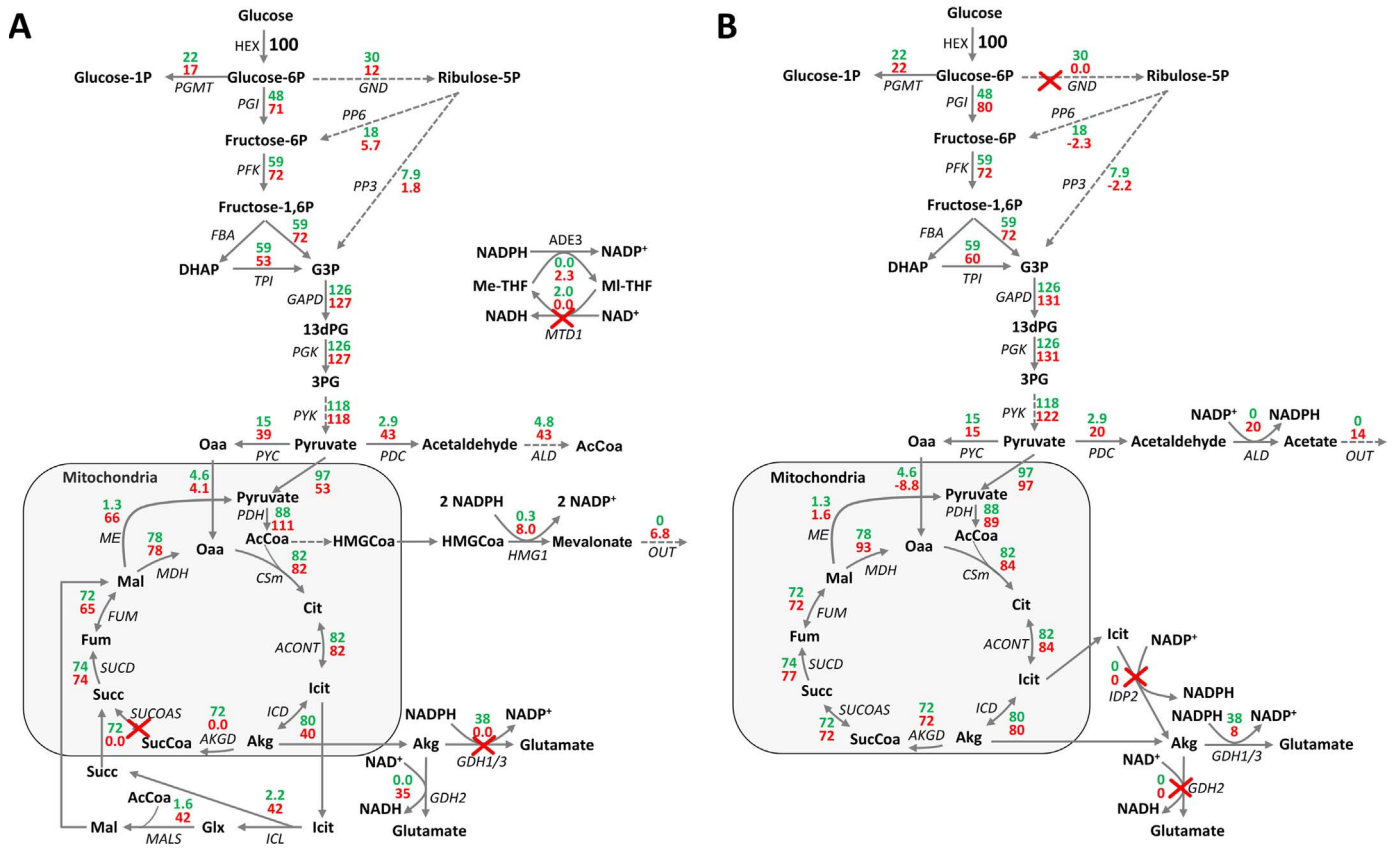

Fig. 4. Flux distributions simulated with lMOMA for the mevalonate mutant (A) and the acetate mutant (B) using the improved flux distribution as the reference. The wild-type flux values are shown in green, the mutant values are shown in red and the knock-outs are signalled by a red cross. The flux values shown are normalized to a glucose uptake rate of 100 arbitrary units. Reactions: ACONT- aconitase, ADE3- cytosolic C1-tetrahydrofolate synthase, AKGD- alpha-ketoglutarate dehydrogenase, ALD- aldehyde dehydrogenase, CSm- citrate synthase, FBA- fructose 1,6-bisphosphate aldolase, FUM- fumarase, GAPD- glyceraldehyde-3-phosphate dehydrogenase, GDH1/3- $NADP^+$-dependent glutamate dehydrogenase, GDH2- NAD^+-dependent glutamate dehydrogenase, GND- 6-phosphogluconate dehydrogenase, HEX- hexokinase, HMG1- HMG-CoA reductase, ICD- mitochondrial isocitrate dehydrogenase, ICL- isocitrate lyase, IDP2- cytosolic NADP-specific isocitrate dehydrogenase, MALS- malate synthase, MDH- mitochondrial malate dehydrogenase, ME- mitochondrial malic enzyme, MTD1- NAD-dependent 5,10-methylenetetrahydrafolate dehydrogenase, PDC- pyruvate decarboxylase, PDH- pyruvate dehydrogenase, PFK- phosphofructokinase, PGI- phosphoglucose isomerase, PGK- 3-phosphoglycerate kinase, PGMT- phosphoglucomutase, PP3- sum of the non-oxidative reactions of the pentose phosphate pathway producing glyceraldehyde-3-phosphate, PP6- sum of the non-oxidative reactions of the pentose phosphate pathway producing fructose-6-phosphate, PYC- pyruvate carboxylase, PYK- pyruvate kinase, SUCD- succinate dehydrogenase, SUCOAS- succinyl-CoA ligase, THRS- threonine synthase, TPI- triose phosphate isomerase; Metabolites: 13dPG- 1,3-diphosphoglycerate, 3PG- 3-phosphoglycerate, AcCoA- acetyl-CoA, Akg- 2-oxoglutarate, Cit- citrate, DHAP- dihydroxyacetone-phosphate, Fum- fumarate, G3P- glyceraldehyde-3-phosphate, Glx- glyoxylate, Icit- isocitrate, HMGCoa- 3-hydroxy-3-methylglutaryl-CoA, Mal - L-malate, Oaa- oxaloacetate, Succ- succinate, SucCoa- succinyl-CoA. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

of required knock-outs also includes the cytosolic NADP-specific isocitrate dehydrogenase (*IDP2*), which can also be a possible source of NADPH for the cell. The final two deletions act on the other side of NADPH metabolism, i.e., on possible sinks of this cofactor. The deletion of the NADH-dependent glutamate dehydrogenase (*GDH2*) is necessary to "force" the use of the NADPH dependent enzyme (GDH1/GDH3) for ammonium assimilation. With the same goal, deleting the cytosolic aspartate aminotransferase (*AAT2*) also increases the flux through the NADPH dependent glutamate dehydrogenase. As shown in Fig. 4, all four deletions trigger rearrangements on the NADPH metabolism that culminate in an increased flux through the cytosolic $NADP^+$-dependent aldehyde dehydrogenase (gene name – *ALD6*, ID in Fig. 4 - ALD). Since this reaction becomes the main source of NADPH in the cytosol, the extra flux results in acetate accumulation.

Experimental data regarding the inactivation of the glucose-6-phosphate dehydrogenase (*ZWF1*) in *S. cerevisiae* has shown that this knock-out can decrease the availability of cytosolic NADPH, resulting in reduced growth on glucose minimal medium (Blank et al., 2005), methionine auxotrophy (Thomas et al., 1991) and increased sensitivity to oxidizing agents (Nogae and Johnston, 1990). In the absence of the glucose-6-phosphate dehydrogenase it has been shown that the cytosolic NADP-specific isocitrate dehydrogenase (*IDP2*) can also fulfil the NADPH production role (Minard and McAlister-Henn, 2005). Furthermore, the overexpression of the cytosolic $NADP^+$-dependent aldehyde dehydrogenase (*ALD6*) has been demonstrated to solve the methionine auxotrophy by increasing the availability of cytosolic NAPDH (Grabowska and Chelstowska, 2003). The compensation between the enzymes mentioned above, as the cytosolic NADPH source, supports the behaviour of the acetate producing mutant in Table 3 and indicates that the model is predicting accurately the NADPH metabolism when the curated flux distribution is used.

4. Conclusions

In conclusion, we observed that the oldest GEM of *S. cerevisiae* (*i*FF708) showed the best prediction of central carbon fluxes. This fact might be explained by extra validation and curation steps performed during its construction (Förster et al., 2003). *i*FF708 is still preferred by many authors in metabolic engineering projects over the more recent GEMs because of its better description of fluxes in the central metabolism (Asadollahi et al., 2009; Bro et al.,

2006; Brochado et al., 2010; Otero et al., 2013).

The results presented here, showed that the flux distributions of the newer GEMs could be significantly improved by imposing flux constraints on reactions involved in the metabolism of NADPH and NADH. The changes applied to cofactor metabolism had the most impact on central metabolic fluxes related to NADPH metabolism, especially in the oxidative PPP. The reactions involved in the production of NADH were not as affected by the imposed constraints because the simulations of the original models were quite good in this regard. The better predictions of NADH metabolism were most likely a consequence of the central role of this metabolite for ATP production in the oxidative phosphorylation.

Since the cofactor constraints proposed here revealed that similar flux distributions can be obtained with more recent models, it is now possible to take advantage of better compartmentalization, elementally balance reactions and increased genome coverage, without sacrificing the quality of fluxes in the central metabolism. This work also showed that the application of GEMs for strain simulation/design using algorithms that require a reference flux distribution should be preceded by a careful validation of the model predictions of the fluxes for the wild-type organism.

Acknowledgements

The authors thank the FCT Strategic Project of UID/BIO/04469/2013 unit, the project RECI/BBB-EBI/0179/2012 (FCOMP-01–0124-FEDER-027462) and the project “BioInd – Biotechnology and Bioengineering for improved Industrial and Agro-Food processes", REF. NORTE-07–0124-FEDER-000028 co-funded by the Programa Operacional Regional do Norte (ON.2 – O Novo Norte), QREN, FEDER. The work Rui Pereira was supported by a PhD grant from FCT (ref. SFRH/BD/51111/2010).

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.meteno.2016.05.002.

References

Alper, H., Jin, Y.-S., Moxley, J.F., Stephanopoulos, G., 2005. Identifying gene targets for the metabolic engineering of lycopene biosynthesis in *Escherichia coli*. Metab. Eng. 7, 155–164. http://dx.doi.org/10.1016/j.ymben.2004.12.003.

Asadollahi, M.A., Maury, J., Patil, K.R., Schalk, M., Clark, A., Nielsen, J., 2009. Enhancing sesquiterpene production in *Saccharomyces cerevisiae* through *in silico* driven metabolic engineering. Metab. Eng. 11, 328–334. http://dx.doi.org/10.1016/j.ymben.2009.07.001.

Avendaño, A., Deluna, A., Olivera, H., Valenzuela, L., Gonzalez, A., 1997. *GDH3* encodes a glutamate dehydrogenase isozyme, a previously unrecognized route for glutamate biosynthesis in *Saccharomyces cerevisiae*. J. Bacteriol. 179, 5594–5597.

Baudry, K., Swain, E., Rahier, A., Germann, M., Batta, A., Rondet, S., Mandala, S., Henry, K., Tint, G.S., Edlind, T., Kurtz, M., Nickels, J.T., 2001. The effect of the erg26–1 mutation on the regulation of lipid metabolism in *Saccharomyces cerevisiae*. J. Biol. Chem. 276, 12702–12711. http://dx.doi.org/10.1074/jbc.M100274200.

Becker, S.A., Feist, A.M., Mo, M.L., Hannum, G., Palsson, B.Ø., Herrgard, M.J., 2007. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. Nat. Protoc. 2, 727–738. http://dx.doi.org/10.1038/nprot.2007.99.

Blank, L.M., Kuepfer, L., Sauer, U., 2005. Large-scale 13C-flux analysis reveals mechanistic principles of metabolic network robustness to null mutations in yeast. Genome Biol. 6, R49. http://dx.doi.org/10.1186/gb-2005–6-6-r49.

Bro, C., Regenberg, B., Förster, J., Nielsen, J., 2006. *In silico* aided metabolic engineering of *Saccharomyces cerevisiae* for improved bioethanol production. Metab. Eng. 8, 102–111. http://dx.doi.org/10.1016/j.ymben.2005.09.007.

Brochado, A.R., Andrejev, S., Maranas, C.D., Patil, K.R., 2012. Impact of stoichiometry representation on simulation of genotype-phenotype relationships in metabolic networks. PLoS Comput. Biol. 8, e1002758. http://dx.doi.org/10.1371/journal.pcbi.1002758.

Brochado, A.R., Matos, C., Møller, B.L., Hansen, J., Mortensen, U.H., Patil, K.R., 2010. Improved vanillin production in baker's yeast through *in silico* design. Microb. Cell Fact. 9, 84. http://dx.doi.org/10.1186/1475–2859-9–84.

Burgard, A.P., Vaidyaraman, S., Maranas, C.D., 2001. Minimal reaction sets for *Escherichia coli* metabolism under different growth requirements and uptake environments. Biotechnol. Prog. 17, 791–797. http://dx.doi.org/10.1021/bp0100880.

Canelas, A.B., van Gulik, W.M., Heijnen, J.J., 2008. Determination of the cytosolic free NAD/NADH ratio in *Saccharomyces cerevisiae* under steady-state and highly dynamic conditions. Biotechnol. Bioeng. 100, 734–743. http://dx.doi.org/10.1002/bit.21813.

Castegna, A., Scarcia, P., Agrimi, G., Palmieri, L., Rottensteiner, H., Spera, I., Germinario, L., Palmieri, F., 2010. Identification and functional characterization of a novel mitochondrial carrier for citrate and oxoglutarate in *Saccharomyces cerevisiae*. J. Biol. Chem. 285, 17359–17370. http://dx.doi.org/10.1074/jbc.M109.097188.

Choi, H.S., Lee, S.Y., Kim, T.Y., Woo, H.M., 2010. *In silico* identification of gene amplification targets for improvement of lycopene production. Appl. Environ. Microbiol. 76, 3097–3105. http://dx.doi.org/10.1128/AEM.00115–10.

Dijken, J.P., Scheffers, W.A., 1986. Redox balances in the metabolism of sugars by yeasts. FEMS Microbiol. Lett. 32, 199–224. http://dx.doi.org/10.1111/j.1574–6968.1986.tb01194.x.

Duarte, N.C., Herrgård, M.J., Palsson, B.Ø., 2004. Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. Genome Res. 14, 1298–1309. http://dx.doi.org/10.1101/gr.2250904.

Dubois, E., Grenson, M., Wiame, J.-M., 1974. The participation of the anabolic glutamate dehydrogenase in the nitrogen catabolite repression of arginase in *Saccharomyces cerevisiae*. Eur. J. Biochem. 48, 603–616. http://dx.doi.org/10.1111/j.1432–1033.1974.tb03803.x.

Edwards, J.S., Palsson, B.O., 2000. The *Escherichia coli* MG1655 *in silico* metabolic genotype: its definition, characteristics, and capabilities. Proc. Natl. Acad. Sci. USA 97, 5528–5533. http://dx.doi.org/10.1073/pnas.97.10.5528.

Feist, A.M., Henry, C.S., Reed, J.L., Krummenacker, M., Joyce, A.R., Karp, P.D., Broadbelt, L.J., Hatzimanikatis, V., Palsson, B.Ø., 2007. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Mol. Syst. Biol. 3, 121. http://dx.doi.org/10.1038/msb4100155.

Fleck, C.B., Brock, M., 2009. Re-characterisation of *Saccharomyces cerevisiae* Ach1p: fungal CoA-transferases are involved in acetic acid detoxification. Fungal Genet. Biol. 46, 473–485. http://dx.doi.org/10.1016/j.fgb.2009.03.004.

Förster, J., Famili, I., Fu, P., Palsson, B.Ø., Nielsen, J., 2003. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. Genome Res. 13, 244–253. http://dx.doi.org/10.1101/gr.234503.

Fowler, Z.L., Gikandi, W.W., Koffas, M.A.G., 2009. Increased malonyl coenzyme a biosynthesis by tuning the *Escherichia coli* metabolic network and its application to flavanone production. Appl. Environ. Microbiol. 75, 5831–5839. http://dx.doi.org/10.1128/AEM.00270–09.

Garcia-Albornoz, M.A., Nielsen, J., 2013. Application of Genome-Scale Metabolic Models in Metabolic Engineering. Ind. Biotechnol. 9, 203–214. http://dx.doi.org/10.1089/ind.2013.0011.

Gombert, A.K., Moreira dos Santos, M., Christensen, B., Nielsen, J., 2001. Network identification and flux quantification in the central metabolism of *Saccharomyces cerevisiae* under different conditions of glucose repression. J. Bacteriol. 183, 1441–1451. http://dx.doi.org/10.1128/JB.183.4.1441–1451.2001.

Grabowska, D., Chelstowska, A., 2003. The ALD6 gene product is indispensable for providing NADPH in yeast cells lacking glucose-6-phosphate dehydrogenase activity. J. Biol. Chem. 278, 13984–13988. http://dx.doi.org/10.1074/jbc.M210076200.

Haselbeck, R.J., McAlister-Henn, L., 1993. Function and expression of yeast mitochondrial NAD- and NADP-specific isocitrate dehydrogenases. J. Biol. Chem. 268, 12116–12122.

Heavner, B.D., Smallbone, K., Price, N.D., Walker, L.P., 2013. Version 6 of the consensus yeast metabolic network refines biochemical coverage and improves model performance. Database (Oxf.). 2013, bat059 . http://dx.doi.org/10.1093/database/bat059.

Herrgård, M.J., Swainston, N., Dobson, P., Dunn, W.B., Arga, K.Y., Arvas, M., Blüthgen, N., Borger, S., Costenoble, R., Heinemann, M., Hucka, M., Le Novère, N., Li, P., Liebermeister, W., Mo, M.L., Oliveira, A.P., Petranovic, D., Pettifer, S., Simeonidis, E., Smallbone, K., Spasić, I., Weichart, D., Brent, R., Broomhead, D.S., Westerhoff, H.V., Kirdar, B., Penttilä, M., Klipp, E., Palsson, B.Ø., Sauer, U., Oliver, S.G., Mendes, P., Nielsen, J., Kell, D.B., 2008. A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. Nat. Biotechnol. 26, 1155–1160. http://dx.doi.org/10.1038/nbt1492.

Jouhten, P., Rintala, E., Huuskonen, A., Tamminen, A., Toivari, M., Wiebe, M., Ruohonen, L., Penttilä, M., Maaheimo, H., 2008. Oxygen dependence of metabolic fluxes and energy generation of *Saccharomyces cerevisiae* CEN.PK113–1A. BMC. Syst. Biol. 2, 60. http://dx.doi.org/10.1186/1752–0509-2–60.

Jung, Y.K., Kim, T.Y., Park, S.J., Lee, S.Y., 2010. Metabolic engineering of *Escherichia coli* for the production of polylactic acid and its copolymers. Biotechnol. Bioeng. 105, 161–171. http://dx.doi.org/10.1002/bit.22548.

Kim, B., Kim, W.J., Kim, D.I., Lee, S.Y., 2014. Applications of genome-scale metabolic network model in metabolic engineering. J. Ind. Microbiol. Biotechnol. . http://dx.doi.org/10.1007/s10295–014-1554–9

Kim, J., Reed, J.L., 2012. RELATCH: relative optimality in metabolic networks

explains robust metabolic and regulatory responses to perturbations. Genome Biol. 13, R78. http://dx.doi.org/10.1186/gb-2012–13-9-r78.
Kim, J., Reed, J.L., Maravelias, C.T., 2011. Large-scale bi-level strain design approaches and mixed-integer programming solution techniques. PLoS One 6, e24162. http://dx.doi.org/10.1371/journal.pone.0024162.
Kingsbury, J.M., McCusker, J.H., 2010. Homoserine toxicity in *Saccharomyces cerevisiae* and *Candida albicans* homoserine kinase (thr1Delta) mutants. Eukaryot. Cell 9, 717–728. http://dx.doi.org/10.1128/EC.00044–10.
Kuepfer, L., Sauer, U., Blank, L.M., 2005. Metabolic functions of duplicate genes in *Saccharomyces cerevisiae*. Genome Res. 15, 1421–1430. http://dx.doi.org/10.1101/gr.3992505.
Leber, R., Landl, K., Zinser, E., Ahorn, H., Spok, A., Kohlwein, S.D., Turnowsky, F., Daum, G., 1998. Dual localization of squalene epoxidase, Erg1p, in yeast reflects a relationship between the endoplasmic reticulum and lipid particles. Mol. Biol. Cell 9, 375–386. http://dx.doi.org/10.1091/mbc.9.2.375.
Lewis, N.E., Hixson, K.K., Conrad, T.M., Lerman, J.A., Charusanti, P., Polpitiya, A.D., Adkins, J.N., Schramm, G., Purvine, S.O., Lopez-Ferrer, D., Weitz, K.K., Eils, R., König, R., Smith, R.D., Palsson, B.Ø., 2010. Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. Mol. Syst. Biol. 6, 390. http://dx.doi.org/10.1038/msb.2010.47.
Lewis, N.E., Nagarajan, H., Palsson, B.O., 2012. Constraining the metabolic genotype-phenotype relationship using a phylogeny of *in silico* methods. Nat. Rev. Microbiol. 10, 291–305. http://dx.doi.org/10.1038/nrmicro2737.
Machado, D., Herrgård, M., 2014. Systematic evaluation of methods for integration of transcriptomic data into constraint-based models of metabolism. PLoS Comput. Biol. 10, e1003580. http://dx.doi.org/10.1371/journal.pcbi.1003580.
Mahadevan, R., Schilling, C.H., 2003. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. Metab. Eng. 5, 264–276. http://dx.doi.org/10.1016/j.ymben.2003.09.002.
Mannhaupt, G., Stucka, R., Pilz, U., Schwarzlose, C., Feldmann, H., 1989. Characterization of the prephenate dehydrogenase-encoding gene, *TYR1*, from *Saccharomyces cerevisiae*. Gene 85, 303–311. http://dx.doi.org/10.1016/0378–1119(89)90422–8.
Mannhaupt, G., van der Linden, G., Vetter, I., Maurer, K., Pilz, U., Planta, R., Feldmann, H., 1990. Analysis of the THR4 region on chromosome III of the yeast *Saccharomyces cerevisiae*. Yeast 6, 353–361. http://dx.doi.org/10.1002/yea.320060408.
McNeil, J.B., McIntosh, E.M., Taylor, B.V., Zhang, F.R., Tang, S., Bognar, A.L., 1994. Cloning and molecular characterization of three genes, including two genes encoding serine hydroxymethyltransferases, whose inactivation is required to render yeast auxotrophic for glycine. J. Biol. Chem. 269, 9155–9165.
Milne, C.B., Kim, P.-J., Eddy, J.A., Price, N.D., 2009. Accomplishments in genome-scale *in silico* modeling for industrial and medical biotechnology. Biotechnol. J. 4, 1653–1670. http://dx.doi.org/10.1002/biot.200900234.
Minard, K.I., McAlister-Henn, L., 2005. Sources of NADPH in yeast vary with carbon source. J. Biol. Chem. 280, 39890–39896. http://dx.doi.org/10.1074/jbc.M509461200.
Mo, M.L., Palsson, B.O., Herrgård, M.J., 2009. Connecting extracellular metabolomic measurements to intracellular flux states in yeast. BMC. Syst. Biol. 3, 37. http://dx.doi.org/10.1186/1752–0509-3–37.
Monschau, N., Stahmann, K.-P., Sahm, H., McNeil, J.B., Bognar, A.L., 2006. Identification of *Saccharomyces cerevisiae GLY1* as a threonine aldolase: a key enzyme in glycine biosynthesis. FEMS Microbiol. Lett. 150, 55–60. http://dx.doi.org/10.1111/j.1574–6968.1997.tb10349.x.
Nissen, T.L., Kielland-Brandt, M.C., Nielsen, J., Villadsen, J., 2000. Optimization of Ethanol Production in *Saccharomyces cerevisiae* by Metabolic Engineering of the Ammonium Assimilation. Metab. Eng. 2, 69–77. http://dx.doi.org/10.1006/mben.1999.0140.
Nogae, I., Johnston, M., 1990. Isolation and characterization of the *ZWF1* gene of *Saccharomyces cerevisiae*, encoding glucose-6-phosphate dehydrogenase. Gene 96, 161–169. http://dx.doi.org/10.1016/0378–1119(90)90248-P.
Nookaew, I., Jewett, M.C., Meechai, A., Thammarongtham, C., Laoteng, K., Cheevadhanarak, S., Nielsen, J., Bhumiratana, S., 2008. The genome-scale metabolic model *i*IN800 of *Saccharomyces cerevisiae* and its validation: a scaffold to query lipid metabolism. BMC Syst. Biol. 2, 71. http://dx.doi.org/10.1186/1752–0509-2–71.
Nookaew, I., Olivares-Hernández, R., Bhumiratana, S., Nielsen, J., 2011. Genome-scale metabolic models of *Saccharomyces cerevisiae*. Methods Mol. Biol. 759, 445–463. http://dx.doi.org/10.1007/978–1-61779–173-4_25.
Orth, J.D., Thiele, I., Palsson, B.Ø., 2010. What is flux balance analysis? Nat. Biotechnol. 28, 245–248. http://dx.doi.org/10.1038/nbt.1614.
Österlund, T., Nookaew, I., Bordel, S., Nielsen, J., 2013. Mapping condition-dependent regulation of metabolism in yeast through genome-scale modeling. BMC Syst. Biol. 7, 36. http://dx.doi.org/10.1186/1752–0509-7–36.
Osterlund, T., Nookaew, I., Nielsen, J., 2012. Fifteen years of large scale metabolic modeling of yeast: developments and impacts. Biotechnol. Adv. 30, 979–988. http://dx.doi.org/10.1016/j.biotechadv.2011.07.021.
Otero, J.M., Cimini, D., Patil, K.R., Poulsen, S.G., Olsson, L., Nielsen, J., 2013. Industrial systems biology of *Saccharomyces cerevisiae* enables novel succinic acid cell factory. PLoS One 8, e54144. http://dx.doi.org/10.1371/journal.pone.0054144.
Palmieri, L., Agrimi, G., Runswick, M.J., Fearnley, I.M., Palmieri, F., Walker, J.E., 2001. Identification in *Saccharomyces cerevisiae* of two isoforms of a novel mitochondrial transporter for 2-oxoadipate and 2-oxoglutarate. J. Biol. Chem. 276, 1916–1922. http://dx.doi.org/10.1074/jbc.M004332200.
Park, J.H., Lee, K.H., Kim, T.Y., Lee, S.Y., 2007. Metabolic engineering of *Escherichia coli* for the production of L-valine based on transcriptome analysis and *in silico* gene knockout simulation. Proc. Natl. Acad. Sci. USA 104, 7797–7802. doi:10.1073/pnas.0702609104.
Patil, K.R., Rocha, I., Förster, J., Nielsen, J., 2005. Evolutionary programming as a platform for *in silico* metabolic engineering. BMC Bioinformatics 6, 308. http://dx.doi.org/10.1186/1471–2105-6–308.
Rocha, I., Maia, P., Evangelista, P., Vilaça, P., Soares, S., Pinto, J.P., Nielsen, J., Patil, K.R., Ferreira, E.C., Rocha, M., 2010. OptFlux: an open-source software platform for *in silico* metabolic engineering. BMC Syst. Biol. 4, 45. http://dx.doi.org/10.1186/1752–0509-4–45.
Rocha, M., Maia, P., Mendes, R., Pinto, J.P., Ferreira, E.C., Nielsen, J., Patil, K.R., Rocha, I., 2008. Natural computation meta-heuristics for the *in silico* optimization of microbial strains. BMC Bioinformatics 9, 499. http://dx.doi.org/10.1186/1471–2105-9–499.
Sánchez, B.J., Nielsen, J., 2015. Genome scale models of yeast: towards standardized evaluation and consistent omic integration. Integr. Biol. 7, 846–858. http://dx.doi.org/10.1039/C5IB00083A.
Satrustegui, J., Bautista, J., Machado, A., 1983. NADPH/NADP+ ratio: regulatory implications in yeast glyoxylic acid cycle. Mol. Cell. Biochem. 51, 123–127. http://dx.doi.org/10.1007/BF00230397.
Scalcinati, G., Partow, S., Siewers, V., Schalk, M., Daviet, L., Nielsen, J., 2012. Combined metabolic engineering of precursor and co-factor supply to increase alpha-santalene production by Saccharomyces cerevisiae. Microb. Cell Fact. 11, 117. http://dx.doi.org/10.1186/1475-2859-11-117.
Segrè, D., Vitkup, D., Church, G.M., 2002. Analysis of optimality in natural and perturbed metabolic networks. Proc. Natl. Acad. Sci. USA 99, 15112–15117. http://dx.doi.org/10.1073/pnas.232349399.
Shlomi, T., Berkman, O., Ruppin, E., 2005. Regulatory on/off minimization of metabolic flux changes after genetic perturbations. Proc. Natl. Acad. Sci. USA 102, 7695–7700. doi:10.1073/pnas.0406346102.
Thomas, D., Cherest, H., Surdin-Kerjan, Y., 1991. Identification of the structural gene for glucose-6-phosphate dehydrogenase in yeast. Inactivation leads to a nutritional requirement for organic sulfur. EMBO J. 10, 547–553.
Voet, D., Voet, J.G., 2011. Introduction to Metabolism, in: Biochemistry. Wiley, pp. 560–562.
Xu, C., Liu, L., Zhang, Z., Jin, D., Qiu, J., Chen, M., 2013. Genome-scale metabolic model in guiding metabolic engineering of microbial improvement. Appl. Microbiol. Biotechnol. 97, 519–539. http://dx.doi.org/10.1007/s00253–012-4543–9.